be identical with the computer readable form of another application of the applicant on file in the Patent and Trademark Office, reference may be made to the other application and computer readable form in lieu of filing a duplicate computer readable form in the new application if the computer readable form in the other application was compliant with all of the requirements of these rules. The new application shall be accompanied by a letter making such reference to the other application and computer readable form, both of which shall be completely identified. In the new application, applicant must also request the use of the compliant computer readable ''Sequence Listing'' that is already on file for the other application and must state that the paper copy of the ''Sequence Listing'' in the new application is identical to the computer readable copy filed for the other application.

(f) In addition to the paper copy required by paragraph (c) of this section and the computer readable form required by paragraph (e) of this section, a statement that the content of the paper and computer readable copies are the same must be submitted with the computer readable form, *e.g.,* a statement that ''the information recorded in computer readable form is identical to the written sequence listing.''

(g) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing under 35 U.S.C. 111(a) or at the time of entering the national stage under 35 U.S.C. 371, applicant will be notified and given a period of time within which to comply with such requirements in order to prevent abandonment of the application. Any submission in reply to a requirement under this paragraph must be accompanied by a statement that the submission includes no new matter.

(h) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing an international application under the Patent Cooperation Treaty (PCT), which application is to be searched by the United States International Searching Authority or examined by the United States International Preliminary Examining Authority, applicant will be sent a notice necessitating compliance with the requirements within a prescribed time period. Any submission in reply to a requirement under this paragraph must be accompanied by a statement that the submission does not include matter which goes beyond the disclosure in the international application as filed. If applicant fails to timely provide the required computer readable form, the United States International Searching Authority shall search only to the extent that a meaningful search can be performed without the computer readable form and the United States International Preliminary Examining Authority shall examine only to the extent that a meaningful examination can be performed without the computer readable form.

[63 FR 29634, June 1, 1998]

§ 1.822 Symbols and format to be used for nucleotide and/or amino acid sequence data.

(a) The symbols and format to be used for nucleotide and/or amino acid sequence data shall conform to the requirements of paragraphs (b) through (e) of this section.

(b) The code for representing the nucleotide and/or amino acid sequence characters shall conform to the code set forth in the tables in WIPO Standard ST.25 (1998), Appendix 2, Tables 1 and 3. This incorporation by reference was approved by the Director of the Federal Register in accordance with 5 U.S.C. 552(a) and 1 CFR part 51. Copies of ST.25 may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies of ST.25 may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202. Copies may also be inspected at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC. No code other than that specified in these sections shall be used in nucleotide and amino acid sequences. A modified base or modified or unusual amino acid may be presented in a given sequence as the corresponding unmodified base or amino acid if the modified base or modified or

unusual amino acid is one of those listed in WIPO Standard ST.25 (1998), Appendix 2, Tables 2 and 4, and the modification is also set forth in the Feature section. Otherwise, each occurrence of a base or amino acid not appearing in WIPO Standard ST.25 (1998), Appendix 2, Tables 1 and 3, shall be listed in a given sequence as ''n'' or ''Xaa,'' respectively, with further information, as appropriate, given in the Feature section, preferably by including one or more feature keys listed in WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6.

(c) *Format representation of nucleotides.* (1) A nucleotide sequence shall be listed using the lower-case letter for representing the one-letter code for the nucleotide bases set forth in WIPO Standard ST.25 (1998), Appendix 2, Table 1.

(2) The bases in a nucleotide sequence (including introns) shall be listed in groups of 10 bases except in the coding parts of the sequence. Leftover bases, fewer than 10 in number, at the end of noncoding parts of a sequence shall be grouped together and separated from adjacent groups of 10 or 3 bases by a space.

(3) The bases in the coding parts of a nucleotide sequence shall be listed as triplets (codons). The amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be typed immediately below the corresponding codons. Where a codon spans an intron, the amino acid symbol shall be typed below the portion of the codon containing two nucleotides.

(4) A nucleotide sequence shall be listed with a maximum of 16 codons or 60 bases per line, with a space provided between each codon or group of 10 bases.

(5) A nucleotide sequence shall be presented, only by a single strand, in the 5 to 3 direction, from left to right.

(6) The enumeration of nucleotide bases shall start at the first base of the sequence with number 1. The enumeration shall be continuous through the whole sequence in the direction 5 to 3. The enumeration shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line.

(7) For those nucleotide sequences that are circular in configuration, the enumeration method set forth in paragraph (c)(6) of this section remains applicable with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant.

(d) *Representation of amino acids.* (1) The amino acids in a protein or peptide sequence shall be listed using the three-letter abbreviation with the first letter as an upper case character, as in WIPO Standard ST.25 (1998), Appendix 2, Table 3.

(2) A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

(3) An amino acid sequence shall be presented in the amino to carboxy direction, from left to right, and the amino and carboxy groups shall not be presented in the sequence.

(4) The enumeration of amino acids may start at the first amino acid of the first mature protein, with the number 1. When presented, the amino acids preceding the mature protein, *e.g.*, pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, shall have negative numbers, counting backwards starting with the amino acid next to number 1. Otherwise, the enumeration of amino acids shall start at the first amino acid at the amino terminal as number 1. It shall be marked below the sequence every 5 amino acids. The enumeration method for amino acid sequences that is set forth in this section remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

(5) An amino acid sequence that contains internal terminator symbols (*e.g.*, ''Ter'', ''*'', or ''.'', etc.) may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences.

(e) A sequence with a gap or gaps shall be presented as a plurality of separate sequences, with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data. A sequence

that is made up of one or more noncontiguous segments of a larger sequence or segments from different sequences shall be presented as a separate sequence.

[63 FR 29635, June 1, 1998]

§ 1.823 Requirements for nucleotide and/or amino acid sequences as part of the application papers.

(a) The ''Sequence Listing'' required by § 1.821(c), setting forth the nucleotide and/or amino acid sequences and associated information in accordance with paragraph (b) of this section, must begin on a new page and must be titled ''Sequence Listing''. The ''Sequence Listing'' preferably should be numbered independently of the numbering of the remainder of the application. Each page of the ''Sequence Listing'' should contain no more than 66 lines and each line should contain no more than 72 characters. A fixed-width font should be used exclusively throughout the ''Sequence Listing.''

(b) The ''Sequence Listing'' shall, except as otherwise indicated, include the actual nucleotide and/or amino acid sequence, the numeric identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the ''Sequence Listing.'' The order and presentation of the items of information in the ''Sequence Listing'' shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an ''M'' is mandatory. The submission of those items of information designated with an ''O'' is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the ''Sequence Listing.'' The following table illustrates the numeric identifiers.

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials.	M.
<120>	Title of Invention		M.
<130>	File Reference	Personal file reference	M when filed prior to assignment of appl. number.
<140>	Current Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if available.
<141>	Current Filing Date	Specify as: yyyy–mm–dd	M, if available.
<150>	Prior Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if applicable include priority documents under 35 USC 119 and 120.
<151>	Prior Application Filing Date.	Specify as: yyyy–mm–dd	M, if applicable.
<160>	Number of SEQ ID NOs.	Count includes total number of SEQ ID NOs.	M.
<170>	Software	Name of software used to create the Sequence Listing.	O.
<210>	SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown.	M.
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues.	M.
<212>	Type	Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be ''DNA.'' In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.	M.
<213>	Organism	Scientific name, i.e. Genus/ species, Unknown or Artificial Sequence. In addition, the ''Unknown'' or ''Artificial Sequence'' organisms shall be further described in the <220> to <223> feature section.	M